(12) INTERNATIONAL APPLICATION PUBLISHED UNDER THE PATENT COOPERATION TREATY (PCT)

(19) World Intellectual Property Organization
International Bureau

(43) International Publication Date
28 November 2002 (28.11.2002)

PCT

(10) International Publication Number
WO 02/095071 A2

(51) International Patent Classification[7]: C12Q 1/68

(21) International Application Number: PCT/NL02/00322

(22) International Filing Date: 22 May 2002 (22.05.2002)

(25) Filing Language: English

(26) Publication Language: English

(30) Priority Data:
01201936.0 22 May 2001 (22.05.2001) EP

(71) Applicant *(for all designated States except US)*: **KONINKLIJKE NEDERLANDSE AKADEMIE VAN WETENSCHAPPEN** [NL/NL]; Kloveniersburgwal 29, NL-1011 JV Amsterdam (NL).

(71) Applicant *(for US only)*: **TIJSTERMAN, Marcel** [NL/NL]; Irislaan 263, NL-2343 CM Oegstgeest (NL).

(72) Inventor; and

(75) Inventor/Applicant *(for US only)*: **PLASTERK, Ronald, H., A.** [NL/NL]; Zwarteweg 5, NL-1405 AA Bussum (NL).

(74) Agent: **PRINS, A.W.**; c/o Vereenigde, Nieuwe Parklaan 97, NL-2587 BN The Hague (NL).

(81) Designated States *(national)*: AE, AG, AL, AM, AT (utility model), AT, AU, AZ, BA, BB, BG, BR, BY, BZ, CA, CH, CN, CO, CR, CU, CZ (utility model), CZ, DE (utility model), DE, DK (utility model), DK, DM, DZ, EC, EE (utility model), EE, ES, FI (utility model), FI, GB, GD, GE, GH, GM, HR, HU, ID, IL, IN, IS, JP, KE, KG, KP, KR, KZ, LC, LK, LR, LS, LT, LU, LV, MA, MD, MG, MK, MN, MW, MX, MZ, NO, NZ, OM, PH, PL, PT, RO, RU, SD, SE, SG, SI, SK (utility model), SK, SL, TJ, TM, TN, TR, TT, TZ, UA, UG, US, UZ, VN, YU, ZA, ZM, ZW.

(84) Designated States *(regional)*: ARIPO patent (GH, GM, KE, LS, MW, MZ, SD, SL, SZ, TZ, UG, ZM, ZW), Eurasian patent (AM, AZ, BY, KG, KZ, MD, RU, TJ, TM), European patent (AT, BE, CH, CY, DE, DK, ES, FI, FR, GB, GR, IE, IT, LU, MC, NL, PT, SE, TR), OAPI patent (BF, BJ, CF, CG, CI, CM, GA, GN, GQ, GW, ML, MR, NE, SN, TD, TG).

Published:
— *without international search report and to be republished upon receipt of that report*

For two-letter codes and other abbreviations, refer to the "Guidance Notes on Codes and Abbreviations" appearing at the beginning of each regular issue of the PCT Gazette.

WO 02/095071 A2

(54) Title: MEANS AND METHODS FOR IDENTIFYING GENES AND PROTEINS INVOLVED IN THE PREVENTION AND/OR REPAIR OF A REPLICATION ERROR

(57) Abstract: The DNA in a cell is prone to mutation. Since too many mutations are detrimental to the survival of a cell or an organism, special mechanism have developed to prevent and/or repair at least part of such mutations. Many of these mechanisms act before the mutation becomes fixed into the genome through replication of the DNA. Some of the genes involved in the prevention of mutations have been identified. However, prior to the invention there was no coherent and systematic way of doing so. The means and methods of the invention enable a person skilled in the art to determine whether a product of a gene is involved in the prevention of a mutation. Identified genes can be used to develop diagnostic tools or used as a target for drug development to manipulate cells on the basis of the presence or absence of function of this gene. Since DNA instability is one of the reasons for rapid tumor progression, it will be useful to provide cancer cells with additional product of such genes for instance through gene therapy.

Title: Means and methods for identifying genes and proteins involved in the prevention and/or repair of a replication error.

The invention relates to the fields of molecular biology and medicine. The invention in particular relates to the identification and use of cellular pathways that are important for maintaining DNA integrity in a cell.

Human tumors arise by multiple mutations that turn so-called proto-oncogenes into active oncogenes, and/or inactivate tumor suppresor genes. Each of these event is the result of a somatic mutation. The chances of getting the "right" combination of mutations to turn a normal cell into a tumor cell are very small, given the inherent stability of the genome. These chances are of course much enhanced if one of the earliest events in the genetic pathway from normal cell to tumor cell is a mutation that enhances the overall level of mutations. Such mutations are called "mutator" mutations.

A simplified calculation to illustrate this: say that 6 mutations are needed within one clonal cell line. Assume that in a mutator cell line the level of mutations is 100 times higher than in a wild type cell. Then the chance of the combination of 6 mutations that make a full blown cancer cell is 100 to the 6th power higher than in a non-mutator cell, or 10 to the 12th power. Such calculations are quite old, and in a sense it could not have been a surprise when it was found that indeed many human cancer cells are mutators.

One common type of mutator genes is DNA mismatch repair. This system recognizes small DNA replication errors, and corrects them. The replication machinery tends to slip on stretches or simple repeat sequences; the resulting repeat instability is also prevented by DNA mismatch repair. Many human tumors are apparently defective in mismatch repair; since one can recognize repeat instability . Indeed in approximately 50% of these tumors one can find a mutation in the known DNA mismatch repair genes (such as

MSH and MLH genes). This confirms that indeed an early event in tumorigenesis is a chance mutation that damages a system that serves to stabilize the genome; then in the resulting unstable genetic background it is much more likely than before that the oncogenic mutations can occur.

These mismatch repair genes were not originally discovered in tumor cells. The known DNA mismatch repair genes were initially discovered in unicellular model organisms (bacteria), as mutator mutants, in which the levels of DNA mutations were enhanced. One case of a hereditary human cancer (HNPCC) was found to be caused by a mutation in a mismatch repair genes, and subsequently one could obviously inspect all the known homologs of factors involved in bacterial mismatch repair for a role in human cancers. But how to get to the other mutator genes? We know that in some classes of tumors 50% of the tumors that show repeat instability do not show a mutation in a known mismatch repair genes, and must thus harbor a mutation in another mutator gene. On top of that there may be mutators that affect mutation levels without showing repeat instability, and thus the actual number of human cancer-causing mutators may be higher than we can now know.

How to get to these genes? Again model organism biology must come to help to indicate candidate genes. These can then be inspected in human tumor samples for possible inactivating mutations. Such candidates if selected from non-human sources ideally fulfill the following criteria:

1. Loss or reduction of function of the gene must result in a significantly enhanced level of mutation rate in the cell lineage.

2. There are homologs in the human genome.

Since animals are in many respect different from bacteria, it may be that at least some levels of genome stabilizing systems are not present in bacteria, but are unique to animals. Therefore these mutators genes are ideally sought in an animal system. On the other hand many factors involved

in DNA metabolism, cell cycle, etc. are very conserved in evolution, so that one may be able to discover relevant genes in simple non-vertebrate model animals.

DNA mismatch repair (MMR) mutants were originally found in screens directed at the identification of bacterial mutants that had a mutator phenotype, and thus had elevated levels of spontaneous mutants in their progeny. Subsequent genetic as well as biochemical studies identified the mismatch repair machinery as an enzymatic complex that could recognize DNA mismatches resulting from single nucleotide substitutions or small insertions/deletions, that could recognize the parental from the newly synthesized strand, excise the new strand around the lesion, and initiate repair to close the gap.

One of the greatest success stories of model organism genetics came when a human syndrome of cancer predisposition, HNPCC for Human Non-Polyposis Colon Cancer, was found to result from a defect in human homologues of genes encoding components in the bacterial mismatch repair machinery (Fishel et al., 1993; Leach et al., 1993; Bronner et al., 1994; Kolodner et al., 1994, 1995; Liu et al., 1994; Nicolaides et al., 1994; Papadopoulos et al., 1994). The fact that these cancers are characterized by an increased instability of simple DNA repeats provided the first clue that a replication-associated repair mechanism was involved (Peinado et al., 1992; Aaltonen et al., 1993; Ionov et al., 1993; Peltomaki et al., 1993). The notion that MMR defects are associated with human cancer provides strong support for the hypothesis that a so-called mutator phenotype, here as a result of elevated levels of unrepaired somatic DNA mismatches, can promote tumorigenesis (Loeb, 1991). This model has been further supported by mouse knockouts of the MMR genes *msh2*, *msh6*, Pms2 or Mlh1 that show enhanced cancer frequencies and repeat instability (de Wind et al., 1995; Reitmar et al.,

1995; Edelmann et al., 1996; Baker et al., 1996; Narayanan et al., 1997; Prolla et al., 1998).

Also in humans that do not contain germline mutations in DNA mismatch repair genes, tumors are often found that display repeat instabilities. Upon analysis these are sometimes defective in known components of the MMR machinery; either they carry mutations within the genes itself or the expression of these MMR genes is epigenetically down-regulated as a result of hypermethylation (Kane et al., 1997; Cunningham et al., 1998; Herman et al., 1998; Veigl et al., 1998). Interestingly not all sporadic human tumors with repeat instability show a defect in the known DNA mismatch repair genes (Liu et al., 1996). In addition, in approximately 30% of HNPCC cases no germline mutations were found in the known MMR genes (Peltomaki and de la Chapelle, 1997; Lynch and Smyrk, 1998). This suggests that there are additional genes in humans but also in other organisms or cells whose loss results in this specific type of genetic instability. These genes can not be easily traced; the currently known genes were only traced based upon prior insights into the mechanism of DNA mismatch repair in model organisms.

In one aspect the invention provides a method for determining whether a product of a gene is involved in preventing a replication error in a cell comprising providing said cell with a specific inhibitor of said product and determining the level of functional expression of a marker gene in said cell, wherein the level of expression of said marker gene is dependent on the occurrence of said replication error. With this method it is not only possible to determine whether a gene is directly involved in preventing a replication error, it is also possible to determine whether a gene influences the efficiency with which the process occurs.

Replication errors usually comprise nucleic acid deletions, nucleic acid insertions and/or base alterations. Replication errors typically occur when

mismatch repair systems fail to correct mutations that occurred between two division cycles. Replication errors can affect the level of functional expression of a marker gene in many different ways. For instance, modification of an enhancer, silencer sequence involved in regulating expression of said maker gene. A replication error can also lead to a change in the coding region of said marker gene whereby said change results a reduction or complete abolishment of the activity of a gene product of said marker gene. Another example of the level expression of said marker gene being dependent on said replication error is the (dis)appearance of an epitope in a gene product of said marker gene as a result of said replication error. Said epitope being detectable with a binding molecule specific for said epitope. Thus many different types of replication errors can influence functional expression of said marker gene.

In a preferred embodiment of the invention said replication error comprises nucleic acid repeat instability. Nucleic acid repeat instability is a replication error that occurs particularly frequent. Several genes have been shown to be involved in preventing nucleic acid repeat instability in a cell. Typical examples are *msh2*, *msh6*, *Pms2* and *Mlh1*. The absence of expression of these genes has been correlated with enhanced cancer frequencies. With a method of the invention it is possible to find additional genes involved in preventing a replication error in a cell. A method of the invention is particularly advantageous for finding additional genes involved in preventing nucleic acid repeat instability in a cell. With the term replication error are not only meant errors that occur during replication. Often, errors occur before replication. Such errors can become fixed in the genome, upon replication of the DNA. The term "replication errors" therefore refers to errors that are introduced into the DNA and that are stable, or stabilized during replication of the cell.

Preventing a replication error in a cell can be done in many ways. Typically, preventing is achieved by preventing a mutation to become fixed in the genome by means of replication of said cell. This can for instance be

achieved by improved repair of mutations such that typically more are corrected prior to fixation. Another method for preventing a mutation error to become fixed in the genome is to (temporarily) inhibit cell division thus allowing for more time in which said mutation can be repaired by the repair machinery of the cell.

For the present invention the phrase "functional expression of a marker gene" is defined as expression of a detectable part of a product of said marker gene. Preferably, activity of a product of said marker gene is detected. However, detection of functional expression can also be done by means of detecting the presence of a particular epitope specific for a gene product of said marker gene. Activity of a promoter or even total amount of marker gene protein may stay essentially the same as long as only one epitope of a product of said marker gene is altered or introduced upon said replication error.

Any method for specifically inhibiting a product of a gene in a cell is suitable for performing the invention. However, a particularly suitable gene specific inhibitor comprises gene specific RNA. Anti-sense RNA, for instance, is very effective in significantly reducing expression of specific genes, particularly in plants cells. Anti-sense RNA can also be made very effective in animal cells. In a preferred embodiment, said specific inhibitor comprises gene specific double stranded RNA. Specific double stranded RNA and particularly RNAi (Fire *et al.*, 1998, Fraser *et al.*, 2000) is very effective in significantly reducing expression of specific genes, also in mammalian cells (Brummelkamp et al, 2002; Elbashir, 2001). In a particularly preferred embodiment said specific inhibitor of a gene product comprises RNAi. A gene specific inhibitor does not necessarily have to be specific for only one gene. A gene specific inhibitor can also be specific for a collection of genes as long as said collection of genes comprises a region of significant homology.

It is possible to use any type of cell in a method of the invention. Culture cells are particularly accessible for manipulation. Moreover, these types of cells can be grown to large numbers thus facilitating detection of

expression of marker genes. However, cell culture cells have a drawback in that many of them already contain unstable genomes. Therefore we prefer to study genome stability in the context of a complete animal. In a preferred embodiment said organism comprises *C. elegans*. *C. elegans* contains a limited number of cells of which the differentiation route and ancestry are completely resolved. In a preferred embodiment said non-human organism is transgenic for said marker gene. In this case it is possible to identify cell type specific genes involved in preventing a replication error in a cell. The method allows one to screen all genes in the C. elegans genome systematically for their possible role in maintaining chromosome stability. We constructed a transgenic animal in which a colorimetrically visualizable gene (lacZ) would only be expressed after a mutation in a short DNA repeat sequence. We confirmed that indeed in such a transgenic animal one could see little patches of blue cells, but only if one had inactivated a known DNA mismatch repair gene (such as MSH, mentioned above). We then found that the same effect can be reached if the MSH gene is inactivated not by mutation but silenced by a phenomenon called RNA interference (RNAi). An advantage of RNAi is that it does not completely knock out gene function in all cells of the body, so that we can detect RNAi effects even if the silenced gene is itself essential for life; in that case RNAi on a population of animals will result perhaps in many early deaths, but in the few escaping animals we find that we can still score the blue pathches that result from the mutator effect (Tijsterman et al, 2002)

Using this method we initially studied all 2000 genes that map on chromosome I of *C. elegans*. Among the genes that we found to have a mutator effect are very plausible candidates, such as the cell cycle checkpoint genes cdc-1 and cdc-5, and the rpa-2 gene, a homolog of gene known to be involved in DNA repair in yeast. But there are also some genes whose function was thus far unknown (see table 3). We have extended this analysis to approximately all 19000 genes that are encoded by the *C. Elegans* genome. Genes found to have a mutator effect are listed in table 4.

We here describe a method that allows one to detect genes that are likely candidates to be the cause of a high proportion of human tumors. Such genes are useful in diagnosis, and treatment choice. Tumors of one mutator type may have a different prognosis or response to a given therapy than another. This can only be tested once these mutator genes are known (and their human counterparts). Such genes are also useful to design new drugs. Of course tumors are only detected once the genetic damage has been done, but still the chances of additional new instability (leading to e.g. escape from drug chemotherapy by mutations in drug resistance genes) will go down upon chemically activate parallel mutator pathways, or by gene-therapy repair or strengthen the damaged mutator gene function. Knowledge of the common mechanisms that cause human cancers aids in defining strategies that protect individuals against such mutator effects, and is thus a way of prevention. Other uses entails life style or dietary advises, food supplements, etc. The invention therefore also provides the use of a mammalian, and preferably human homologue of a gene obtainable by a method of the invention in a method for diagnosis, prognosis, gene therapy and drug targeting approaches.

Any gene can be a marker gene provided that a product of said gene can be detected. Expression of said marker gene, and particularly changes in the expression level of said marker gene must be detectable. Preferably, said marker gene is not performing a critical function in said cell. Preferably, said marker gene is provided to said cell. Suitable marker genes are LacZ and GFP, although other equally suited marker genes are readily available. In a preferred embodiment said marker gene comprises LacZ.

Many types of replication errors can result in a change in the level of expression of a marker gene in a cell. In a preferred embodiment said replication error comprises an error that results in a frame-shift in a protein

coding domain of said marker gene. In a particularly preferred embodiment said replication error comprises a deletion/insertion in or of a mono- or di-nucleotide repeat and wherein said deletion and/or insertion results in a frame-shift in or of said protein coding domain, wherein said frame-shift result in a change in the level of functional expression of said marker gene. In a preferred embodiment said frame shift results in a functional protein, preferably an easily detectable function that is not critical to the cell. Detection of said function can subsequently be used to measure the level of functional expression of said marker gene. Preferably, said frame-shift results in functional LacZ or GFP expression.

In one aspect the invention provides a method of the invention further comprising identifying said gene involved in preventing nucleic acid repeat instability in a cell. Once identified it is of course very easy to isolated said gene through known methods in the art. It is even possible to synthetically generate said gene. The invention therefore also provides an isolated and/or recombinant gene obtainable by a method according to the invention. In a preferred embodiment said isolated and/or recombinant gene comprises a sequence as listed in table 3 or table 4 or an equivalent thereof. An equivalent of a gene as listed in table 3 or table 4 is preferably a human homologue thereof. A significant fraction of human tumors is apparently caused by somatic mutations in genes that affect genome stability, but not nearly in all cases these mutations are in genes of the known mismatch repair system. There seems no direct way to identify these genes, while they may be highly relevant as causative agents of human cancers. An aspect of the present invention provides a system that mimics the somatic repeat stability in human cancers. With the means and methods of the invention it is possible to determine whether a product of a gene is involved in preventing a replication error in a cell. It is further possible to identify the product and the gene. Identified genes can be isolated and/or cloned. Such isolated and/or

recombinant genes can further be used in a large variety of methods known to the person skilled in the art. In a preferred embodiment the invention provides a method for determining whether a cell is predisposed to display a nucleic acid repeat instability phenotype comprising determining functional expression of a gene according to the invention in said cell or derivative thereof. Preferably said gene is a gene as listed in table 3 or table 4 or an equivalent thereof. Preferably, said equivalent is a human homologue of a gene listed in table 3 or table 4. Human homologues may be found by sequence comparison. Human homologues may also be found based on a function of the proteins in the two species. A homologue of gene identified in a method of the present invention, comprises a similar function in kind in another species, not necessarily in amount, as the gene identified with a method of the invention. A nucleic acid repeat instability phenotype is for instance cancer, or an immune deficiency. The method may be performed through any means for determining whether a gene is expressed in a functional way. One way is to determine whether said gene is intact in said cell. Typically this is done on a nucleic acid sequence level. Alternatively, expression levels can be detected by means of for instance an antibody specific for a proteinaceous product of said gene in said cell or a method for detection of RNA. In a preferred embodiment said cell is present in a clinical sample. In this way it may be determined whether an individual is predisposed of developing a disease associated with instability of the genome. The method can therefore advantageously be used to determine whether an individual is predisposed to display a nucleic acid repeat instability phenotype. However, diagnostic tools of the invention may also be used, alone or in combination with other methods, to determine whether said cell is a cancer cell, or predisposed to become a cancer cell and which type of mutator mutations is responsible for its etiology (which may play a role in prognosis, therapy choice and possibly in therapy development. Said cell may of course also be part of, or be derived from a non-human organism. In this

way, individuals may be found, or screened for that comprise alterations in the functional expression of said gene.

The invention further provides a kit for performing a method for determining whether a cell is predisposed to display a nucleic acid repeat instability phenotype, said kit comprising a means for determining functional expression of a gene identifiable with a method of the invention Preferably, said kit comprises an antibody specific for a gene product of a gene identifiable with a method according to the invention. In a preferred embodiment said kit comprises a probe for a gene identifiable with a method of the invention or a probe for a gene product of said gene. In yet another aspect said kit comprises means for obtaining at least a functional part of sequence of a gene identifiable with a method according to invention, or a functional part of a sequence of a gene product of said gene. A functional part of a sequence comprises at least a part sufficient for the identification of said gene (gene product) and/or the determination whether said gene and/or a product derived from it comprises an alteration such that its activity in preventing a replication error in a cell is modified and preferably decreased. Typically, a functional part comprises at least 20 nucleotides or 7 amino-acids.

The invention provides means and methods for identifying genes and gene products involved in preventing a replication error in a cell. With the tools provided by the invention, it is possible to identify essentially all genes and/or gene products involved in the prevention of a replication error in a cell. The identification aspect of the invention is exemplified below for *C. elegans*. Of course this is just one way of obtaining the desired result. Most research on mismatch repair function *in vivo* has focused either on unicellular organisms such as bacteria or yeast, because in those one can easily monitor mutator effects in large numbers of progeny, or in somatic cells or tissue culture cells of higher animals. The numbers of progeny animals that need to be inspected to

recognize spontaneous mutants (that are not induced by chemicals or radiation) is prohibitively large. It was therefore possibly to be expected but not established that the mismatch repair machinery contributed significantly to removal of point mutations from progeny in multicellular organisms. It would not have been impossible that the mismatch repair system acted only to protect against base pair substitutions that arise in somatic cells. However, we find that the mismatch repair system in a metazoan animal such as *C. elegans* has pretty much the same effect on progeny that it has on that of unicellular organisms: a protection of a factor 20 against mutations, most of which are transitions and frameshifts.

This protection is as important for the male germline as for the female (actually hermaphrodite) oocytes. Note that we did not address the role of DNA mismatch repair in hermaphrodite sperm, since experimentally the mutations that arise in self-fertilizing hermaphrodites can not easily be attributed to the sperm or the oocytes.

Genes capable of preventing a replication error in a *C. Elegans* cell can be used to screen for homologues of said gene in other organisms. It is likely that such a homologue will also have the property of preventing a replication error in a cell of that organism. A person skilled in the art is well capable of verifying this property in a homologue. Particularly preferred homologues are of course human homologues.

The level of spontaneous mutagenesis in the *msh*-6 mutant strain per generation is 10 fold lower than that induced by the most efficient chemical mutagens. Therefore it is not surprising that one recognizes different visible mutants among progeny of *msh*-6 animals; since the mutator effect is continuous, one could in principle culture the strain for multiple generations and achieve quite significant accumulated levels of mutations (while maintaining selection pressure for viability). Possibly a strain like this may be of use in experiments aimed experimental quantitative genetics, where genetic

adaptation to specific environmental challenges can be studied more efficiently that in a wild type isolate, because the rate of evolution is enhanced.

One of the most spectacular aspects of RNA interference is that it also works when *C. elegans* is fed on dsRNA or even on E. coli strains that are genetically modified to produce *C. elegans* specific dsRNAs (Timmons and Fire 1999). Thus far these effects were always transient, and did not persist longer than two or three generations, when apparently the RNAi machinery had been diluted out. Since we here study a gene whose function is to protect the genome against mutations, we found that a single episode of exposure to dsRNA was sufficient to induce permanent mutations in the progeny of exposed animals. Fortunately, for higher animals than these small worms there is no evidence that ingested nucleic acids can affect the germline. Since the effect can also be induced by feeding dsRNA for the mismatch repair genes, we now have a system to test any *C. elegans* gene for its role in repressing repeat length changes. Recently genome-wide libraries of dsRNAs of *C. elegans* have been described (Fraser *et al.*, 2000), and we are now testing all genes in this animal's genome for their mutator effect. If additional classes of mutator genes exist, possibly not at all related to mismatch repair, but perhaps to replication factors, chromatin proteins that protect the genome, or totally novel protection systems, they can now be discovered, and possible human homologs can be tested for their role in human cancer etiology.

Now that the invention provides means and methods for determining whether a cell is disposed to display a replication error, it is possible to devise means and methods that capitalize on this capability. In one aspect the invention provides a method for determining whether a compound is capable of influencing a process involved in preventing a replication error in a cell comprising providing said cell with said compound and determining the

level of expression of a marker gene in said cell, wherein the level expression of said marker gene is dependent on said replication error. Preferably, said level is dependent on the occurrence of said replication error. In a preferred embodiment said compound is provided to a collection of said cells. A compound is said to influence the process when the compound reduces or increases the frequency with which a replication error is detected. In a preferred embodiment said method further comprises providing said cell with a specific inhibitor for the expression of a gene involved in preventing a replication error in a cell. In this way the detection of compounds capable of decreasing said frequency is enhanced. Preferably, said gene is a gene obtainable by a method of the invention.

In yet another aspect the invention provides a gene delivery vehicle comprising a gene of the invention or a functional part, derivative and/or analogue thereof. Such a functional part, derivative and/or analogue comprises the same nucleic acid repeat instability preventing activity as said gene in kind, not necessarily in amount. The invention further provides a method for influencing a process involved in preventing a replication error in a cell comprising providing said cell with a gene delivery vehicle of the invention. In this way said cell can be provided with an improved capacity to prevent nucleic acid repeat instability. In one aspect the invention therefore provides the use of a gene delivery vehicle of the invention for the preparation of a medicament.

As used herein the term gene refers to a protein coding domain, it may or may not be accompanied by with local elements that *in cis* regulate expression of said gene. Typical *in cis* elements are promoters, transcription terminator elements, introns and the like. A product of a gene can be a trancribed RNA and/or a translated proteinaceous molecule. With the current technology it is of course possible to generate synthetic versions of each of such

RNA or proteinaceous molecules. Such synthetic versions are of course equivalents of these molecules.

In yet another aspect the invention provides a non-human animal comprising a marker gene wherein the level of expression of said marker gene is dependent on the occurrence of said replication error. Such an animal can favorably used in a method of the invention. Preferably, said marker gene is provided to cells of said animal. In a particular preferred embodiment said animal is transgenic for said marker gene. The invention also provides a method for determining whether a compound is capable of inducing a replication error comprising providing a non-human animal according to any the invention, with said compound and determining in said animal or progeny thereof whether the expression level of said marker gene is altered. Preferably, said non-human animal comprises *C. elegans*.

Said compound can be any compound. In case said compound comprises RNAi than it is possible to study whether said RNAi is capable of inducing a replication error. When said RNAi is designed to be a specific inhibitor for a gene product of a gene from said animal than the method resembles methods that are also described above. When no specific designing is done than it is still possible to study the capability of said RNAi to induce a replication error.

In another embodiment said compound comprises a free radical or a substance capable of generating a free radical, either alone or in combination with another molecule. In general this method is suited to determine and identify compounds that are capable of inducing replication errors in whole organisms.

The invention further provides a method for typing a cell comprising determining in a sample comprising said cell functional expression of a gene listed in table 3 or table 4 and comparing said functional expression with a reference sample.

Examples

Example 1.

Materials and Methods

Strains and maintenance

General methods for culturing *C. elegans* strains were as described in Brenner (1974). Strains used in this study were: CB1500 (*unc-93(e1500)*), MT765 (*unc-93(e1500 n224)*), BC1958 (*dpy-18(e364)/eT1 III; unc-46(e177)/eT1* V). A deletion mutant of *msh*-6: *pk2504* was isolated from a chemical deletion library as described (Jansen *et al.*, 1997).

Spontaneous mutation frequency

Growing cultures of *msh*-6 strains segregate a plethora of visible mutants indicative of a mutator phenotype. From the brood of 4 *msh*-6 hermaphrodites, 300 progeny animals were picked that had a wild type appearance. These worms were grown individually and the progeny was inspected for mendelian segregation of visible phenotypes. Plates were screened a second time 2 days after food deprivation; this allows the scoring of an embryonic lethal phenotype, here interpreted as the abundant presence of dead eggs on the culture dish.

To determine whether *msh-6* animals have a high incidence of male (him) phenotype, the broods of 3-5 animals of genotype *msh-6* or wildtype were

inspected for the presence of a male. *msh-6*: 1/1209 (0.08%), wildtype: 1/1059 (0.09%). The genetic recombination frequency was analyzed by determining the genetic distance between the visible marker *unc-32* and *dpy-18* on LGIII in an *msh-6* and wildtype genetic background. For animals of genotype: *msh-6*; unc-32 dpy-18/+ +, the brood consisted of 412 wildtype, 20 Unc, 21 Dpy, and 112 Unc Dpy: resulting in a recombination frequency of 0.075 (Map distance: 7.5 cM). In a mismatch proficient genetic background the frequency was 0.072 (Map distance: 7.2 cM): 527 wildtype, 26 Unc, 24 Dpy, and 140 Unc Dpy segregated from animals of genotype unc-32 dpy-18/+ +.

The mutator phenotype of *msh*-6 C.elegans was quantified using the reciprocal translocation *eT1 (III;V)* as a balancer, as described by Rosenbluth (1983). First, *msh*-6 males were crossed with hermaphrodites that were homozygous for the translocated *eT1* chromosomes (this genotype results in a visible phenotype because the translocation disrupts the *unc-36* locus). F1 males were subsequently crossed with hermaphrodites of genotype*: dpy-18; unc-46*, (in order to mark the non-translocated chromosomes) and cross progeny of genotype *msh*-6/+ I; *dpy-18/eT1* III*; unc-46/eT1* V were selected. Next generation animals homozygous for *msh*-6 and segregating both Dpy-18 Unc-46 and Unc-36 animals were used as starting strains in the following experimental setup: Phenotypically wild type progeny of hermaphrodites of the above described genotype were picked onto individual plates and scored for segregation of the Dpy-18 Unc-46 phenotype. The frequency of recessive lethal mutations induced in the balanced area of the genome is reflected by the percentage of animals that fail to segregate this phenotype: a lethal in the crossover-suppressed region of the canonical chromosomes prevent embryos homozygous for these chromosomes to developing into adult Dpy Unc worms. Clonal lines that were positive in this screening were continued to grow and confirmed as carrying a lethal mutation inside one of the crossover-suppressed regions if showing no such Dpy Unc in at least 250 offspring.

For determining the germline frequency in male sperm of *msh*-6 animals, males of genotype *msh*-6 I; *dpy-18/eT1* III; *unc-46/eT1* V were crossed to hermaphrodites of genotype *eT1(III/V)*. Phenotypically wild type progeny were analysed for segregation of the marked chromosomes as described above. The germline mutation frequency of hermaphrodite oocytes was determined by analyzing the phenotypically wildtype crossprogeny of *dpy-18/eT1; unc-46/eT1 and eT1* males crossed to *msh*-6; *dpy-18; unc-46* hermaphrodites. In the three crossing schemes, the *msh-6* deficient animals, that were used to start the analysis with, were homozygous for more than one generation. Therefore, in order to prevent scoring mutations that occurred in earlier generations (that are to result in so-called "Jackpots") more than 30 cross-progeny animals were tested from a single hermaphrodite.

RNAi of *msh*-6 and *msh*-2 was done by injecting hermaphrodites of strain BC1958 with cognate dsRNA and subsequent analysis of the mutator phenotype in the phenotypically wildtype F1; Thus the F2 was inspected for segregation of the Dpy Unc phenotype. In addition, RNAi was measured by culturing BC1958 animals on *msh*-2 or *msh*-6 dsRNA producing bacteria (described below).

Mutation spectrum of *msh*-6 worms

Phenotypic reversion of the uncoordinated "rubber-band", egg-laying-defective phenotype conferred by *unc-93(e1500)* was used to determine the nature of mutations that occurred in a *msh-6* genetic background. Cultures started with a single hermaphrodites of genotype *msh-6 unc-93(e1500)* were inspected regularly for revertants that were recognized by their wildtype movement and normal egg-laying behavior. Intragenic reversion events (mutations in at least 4 other loci can suppress the *unc-93(e1500)* associated phenotype) were

identified by the failure of these alleles to complement *unc-93(e1500n224)*. Subsequently, the coding region of the *unc-93* locus was sequenced.

Microsatellite repeat instability in *msh-6* worms

From a single hermaphrodite (*msh-6* and Bristol N2), 55 progeny were picked to start lines that were maintained by transferring several L4 animals every 3-4 days to new plates. After 10 generations DNA was isolated from cultures started with a single animal (due to the mutator phenotype of *msh-6,* mutations will accumulate and often a sterile phenotype is observed when individual animals are cloned out). From these cultures different genomic loci were analyzed by sequencing PCR products. Primers used are (5'-3'): R03C1_A: cggcaaacaatttttccg, R03C_C: acggaggtgttcacggag, F59A3_A: cgtttgaaggatgatgtc, F59A3_C: gatgctcgatgacttcgg, C41D7_A: gattctcaagtccacccg, C41D7_C: gacccgttctcctactcc, M03F4_A: cgaaatggatctgagtggg, M03F4_C: atatcccatgatgacccc, C24A3_A: gagtgcgcttgaagagactg, C24A3_C: cggaactcggagagagatag, Y54G11A_A: ggatcttggctcctggaacg, Y54G11A_C: cattgagtgatactcggccg.

Detection of somatic repeat instability

To allow detection of somatic repeat instability we created several constructs that contained stretches of either mono- or dinucleotde repeats between the start of translation and the *lacZ* ORF, under the control of a heat-shock promoter.

In brief: vector L2681 (Fire-kit), that has a GFP/LacZ fusion under the control of a heat shock promoter, was digested with BamHI and allowed to close to

create pRP1820; this cloning step removes two upstream ATG sequences without affecting essential promoter sequences. Then, the original starting codon was removed by site-directed mutagenesis to create pRP1821. This construct was subsequently used as a recipient for insertion of DNA fragments containing different types of repeats: Partially complementing oligonucleotides were annealed and inserted into a KpnI site near the beginning of the fusion protein encoded sequences. All constructs had a similar molecular architecture: Heat-shock promoter-(KpnI-)-ATG-(A or CA)n-GFP/*LacZ* ORF, (sequences and cloning details available upon request). The different types of repeat used in this study were pRP1822: (A)17, pRP1823: (A) 16, pRP1840: (A) 15, pRP1841: (CA) 15 , pRP1842: (CA) 14, pRP1843 (CA) 13. pRP1823 and pRP1842 contain an in frame LacZ construct encoding functional β-galactosidase.

All constructs were injected separately (together with pRF4 containing the dominant marker *rol-6*) into the canonical *C. elegans* strain BristolN2 to established transgenic lines (Mello et al., 1991). The transgenic array containing pRP1822 was integrated by γ-irradiation and used for further detailed analysis of somatic reversion events.

To identify expression of β-galactosidase, nematodes were fixed and stained with X-gal (5-bromo-4-chloro-3-indolyl-β-D-galactpyranoside). Animals were examined with Nomarski optics.

cDNA analysis

Primarily based on sequence homology comparison with other eukaryotic *msh-6* genes we suspected the GENEFINDER prediction of the *C. elegans msh-6* coding sequence, Y47G6A.11, as annotated in the *C. elegans* database AceDB, to be incorrect. While the N-terminal part of the predicted protein (encoded by Y47G6A.11 exon1 to 7) does not show any significant homology with *msh*-6 orthologs, amino acids encoded by exon-8 are homologues to the N-terminal part of the human protein. In favor, exon-8 predicts an ATG at +1 from a

perfect SL1 splice site. SL1 splicing directly onto the putative exon-8, hereafter named exon1 was confirmed by sequencing DNA material obtained from PCR on cDNA derived from Bristol N2 with primers corresponding to SL1 and *msh*-6 sequences. In addition, we were unable to amplify cDNA with primers directed against the putative upstream exons and exon8 or 9.

RNAi

By injection: PCR fragments of *msh*-6 and *msh*-2 coding sequences were cloned into vector pCCM114 (kind gift of Craig Mello) that contains oppositely oriented T7 promoters. Plasmid DNA was isolated, linearized and used as template to synthesize dsRNA *in vitro* with T7 RNA polymerase (Boehringer Mannheim) according to the manufacturers conditions. Hermaphrodites were injected with 500 ng/µl dsRNA.

By feeding: *msh*-6 and *msh*-2 DNA segments were cloned into the "feeding vector": L4440, subsequently transformed to HT115 bacterial cells that were used for RNAi by feeding using the protocol described by Ahringer and coworkers (Fraser et al., 2000).

A library of bacterial clones, derived from the laboratory of Julie Ahringer (Welcome CRC, UK), that contains all *C. elegans* open reading frames was used to assay individual clones for their potential to induce replication errors, visualized by the detection of somatic repeat instability. To this end individual animals that contain construct pRP1822 were placed on AGAR plates that were seeded with HT115 bacteria; each plate with a different clone and thus expressing RNA of a different *C. elegans* ORF. The next generation *C. elegans* animals were assayed for expression of beta-galactosidase indicative of frameshift errors that occurred in the transgene during development.

Screening of the complete genome of *C. elegans*:

Bacterial clones (HT110) that contain a plasmid, each carrying a specific DNA-insert corresponding to a unique part of a *C. elegans* gene are seeded on standard assay plates as described in Fraser et al.,(2000). The worms are grown for one or two generation, harvested, and assayed for LacZ expression as described above and in Tijsterman et al., (2002). If animals score "positive" for this assay, (that a significant level of expression is noticed), the assay is repeated in 6 fold with the cognate bacterial clone. Bacterial clones that are validated like this are considered to contain DNA sequence corresponding to genes that when knocked down by RNA interference lead to DNA instability. The genes corresponding to these DNA sequences are listed in table 3 and 4. Because the bacterial clones are derived from a library of bacterial clones that was constructed for purposes as described here, the DNA sequence of the clones that are tested are known and kept in a data-base (See Fraser at al.,(2000) for a detailed description of this library).

Results

Mutator phenotype in mismatch repair defective *C. elegans*

We screened the genome sequence of *C. elegans* for homologs of bacterial and human DNA mismatch repair genes, and found *msh-2* and *msh-6* (homologues to the bacterial mutS gene) and *mlh-1* and *pms-1* (homologues to prokaryotic mutL). Surprisingly, an orthologue of *msh-3* was not detected. We then knocked out the *msh-6* gene, using the mutant library approach previously developed in our laboratory (Jansen et al., 1999). Figure 1 shows the human and *S. cerevisiae* homologs aligned to *msh-6* of *C. elegans*, and the deletion mutant that was used in this study.

Homozygous *msh-6* mutants are viable, and the first indication of the mutator phenotype was the frequent occurrence of readily recognizable mutants (Dpy, Unc) among the progeny. Since *C. elegans* lines can be maintained as self-fertilizing hermaphrodites, spontaneous new mutations can homozygose in self-progeny, so that recessive mutations are easily observed. (At least 20 phenotypic mutations were found in 300 progeny of 2 phenotypically wild type hermaphrodites). In the parental strain such level of spontaneous mutations is not seen. To quantify this mutator phenotype, we scored for lethal mutations in a region of the genome that can be genetically monitored (see methods section). In a wildtype strain we detect spontaneous mutations in this region below a frequency of 10^{-3}, which is in line with the numbers reported in the literature (Rosenbluth et al.,1983). In *msh-6* mutants this level is at least a 25 fold elevated (figure 2). Apart from the increased mutation frequency in the *msh-6* mutant, no other phenotype that are indicative for specific defects in genome stability were noticed: X-chromosomal non-disjunction is not affected by the *msh-6* deletion, indicated by the absence of a high incidence of male (him) phenotype. Also, no effect was observed on genetic recombination: the genetic distance between visible markers is similar in wildtype and *msh-6* animals (see materials and methods for details).

These mutations could theoretically arise from mutations that occur uniquely in the sperm or in the oocytes of the hermaphroditic parent. To test whether the mismatch repair machinery protects the male as well as the female germline equally, we performed experiments that scored for spontaneous mutants in progeny from crosses between males and hermaphrodites, in which either one of the parents was mutant and the other wildtype for *msh-6* (see methods for details). As shown in figure 2, both the oocytes of the hermaphroditic mother and the sperm from male fathers show a similar increase in the level of spontenous mutagenesis in the *msh-6* mutant. We conclude two things: the frequency of original DNA replication errors is

probably comparable in sperm and oocytes, and the level of protection by the mismatch repair machinery is also similar.

As a second measure of mutation rates we took the frequency of loss-of-function mutations in the *unc-93*(e1500) mutation. The e1500 allele makes animals hypercontracted, while complete loss of the unc-93 gene has no strong visible phenotype, and thus mutants of the *unc-93*(e1500) gene can be scored by recognizing normally moving animals among contracted ones. Therefore this gene has been previously used to assay mutagenesis levels. We found that the levels of mutations in unc-93(e1500) go up 30 fold in msh-6 mutants compared to wildtype.

The advantage of using the unc-93 monitor gene is that once obtained these mutants can also be identified at the molecular level by direct sequencing of the relatively small genomic *unc-93* gene. It is know that loss of four other genes (*sup-9, sup-10, sup-11* and *sup-18*) also revert the unc-93(e1500) phenotype, so we first sorted out the mutations that mapped to *unc-93*, and sequenced only those. The nature of the mutations is shown in table 1: mostly we find G to A transitions and frameshifts in short monomeric runs, which is similar to the spectrum seen in bacteria, yeast and mammalian tissue culture cells . Note that nothing is know about point mutations in progeny of mismatch repair deficient humans of animals, so that this is the first indication of spontaneous mutation spectra in progeny of repair deficient animals.

Microsatellite instability is a hallmark of tumors derived from HNPCC patients. To see if and to what extend worms defective for *msh*-6 display microsatellite instability we started 50 parallel lines by cloning the progeny of one *msh-6* hermaphrodite. After these lines were maintained for 10 generations we picked one animal per line and sequenced various genomic loci

containing microsatellites. As shown in table 2, especially dinucleotide repeats become highly instable in the absence of functional *msh-6*.

Having observed these fairly frequent repeat length changes in the germline of *msh*-6 mutants we wondered if these could also be observed in somatic cells. With worm living only two weeks, and most somatic cells being only a few cell divisions removed from the zygote, one may not expect too many mutations. Therefore we devised a sensitive system for scoring repeat length instability. We wanted to clone a repeat into a reporter gene, in such a way that the repeat was between the ATG initiation triplet and the domain of the gene encoding the enzymatic activity, and would keep the latter out of frame. Then unrepaired replication errors in the repeat could bring the gene in frame, which could be visualized (see also figure 3). To enhance the chances of finding such events, we could take advantage of the fact that transgenes in *C. elegans* are usually tandem repeats of hundreds of copies of the injected DNA; one would hope that a frame-shift in only one of those copies could be scored. Initial attempts to use GFP for this purpose failed (presumably because the signal of one in-frame GFP gene copy among hundreds of out-of-frame copies was too low). We then constructed a similar plasmid but now using the LacZ reporter (figure 3). A disadvantage of this reporter is that the animal needs to be impregnated with the reagent X-gal, which kills the animal. An advantage is that LacZ staining can be more sensitive, especially because one can prolong the staining to get more signal.

Figure 4 shows staining of transgenic worms after the LacZ transgene is expressed by induction of the heat shock promoter. In the wildtype worms there is virtually no staining. The low level that is seen may reflect a low level of repeat instability even in the wildtype, or it my reflect frameshift errors that are made during translation or both . In *msh*-6 mutant worms, on the other hand, the effect is dramatic almost every worm shows one or more blue patches. We conclude that these arise from repeat instability and restoration of the LacZ reading frame in lineages. Unfortunately the fixated and stained

worms have not allowed us to recognize specific sublineages, but we see blue patches of multiple tissues.

To check the role of the repeat in this *msh*-6 dependent frame shift, we generated transgenic animals that contained identical constructs without the repeat and saw no animals displaying the blue patched pheenotype indicating that the repeat is an essential component of the detection system.

Destabilizing the germline by feeding *msh*-2 and *msh*-6 dsRNA

RNA interference is the silencing of gene expression by administration of dsRNA that corresponds to exonic sequences of that gene (Fire et al., 1998). The most striking effect is that dsRNA can be administered by soaking the worms in it (Tabara et al., 1999), or even by feeding them on E. coli that contains a plasmid that transcribes both strands which can form dsRNA together (Timmons and Fire). We fed worms on E. coli that contained dsRNA for *msh*-6, and measured spontaneous mutation rates by scoring for mutants I the progeny. The results are shown in figure 1: the RNAi effect is comparable to that of a genetic knock-out of *msh*-6

Destabilizing the genetic contents of somatic cells by feeding *msh*-2 and *msh*-6 dsRNA

Combining the somatic repeat stability assay with *msh*-2 and *msh*-6 RNAi, we fed dsRNA to worms, and scored for repeat length changes in somatic cells. As shown in figure 5 the effect is the same as that of the genetic null: almost every animal has LacZ+ patches. This means that the stability of an animals genome is directly influenced by the genetic material it eats.

Brief description of drawings

Figure 1. The *C. elegans msh-6* gene. (A) Structure of the *C. elegans msh-6* gene deduced from genomic sequences and cDNA generated by RT-PCR from Bristol N2 RNA. The genomic region that is deleted in *pk2504* (nt. 24180 - 25956 of Y47G6A, GenBank accession number: AC024791), and takes out exon 5 and part of exon 6 is indicated. (B) Alignment of the amino acid sequence of *C. elegans*, Human and *S. cerevisae* MSH-6 using the CLUSTALW algorithm. Black shading indicates amino acid identity, grey shading indicates conserved amino acid substitutions. The amino acids deleted in *pk2504* are underlined. Possible alternative splicing of exon 4 on to exon 7 predicts an out of frame product.

Figure 2. Mismatch repair proteins MSH-6 and MSH-2 protect the *C. elegans* germline from spontaneous mutagenesis. The experimental setup that is used to measure the level of spontaneous mutagenesis is described in the materials and method section. This assay determines the absolute number of loss of function mutations in essential genes in a region that covers approximately 7% of the *C. elegans* genome (estimated number of target genes: ~300). The y-axis reflects the percentage of animals that acquire such a lethal mutation within one generation.

Figure 3. Outline of the principle to detect somatic repeat instability.

Figure 4. Genetic instability in MMR defective somatic cells. A schematic representation of the constructs that are used to measure somatic repeat instability is depicted above the images of the nematodes. (A) Transgenic *C. elegans* that carry multiple "in-frame" copies of heat shock driven *LacZ*. (B) MMR proficient transgenic *C. elegans* (N2) that carry multiple copies of a *LacZ* containing construct in which a repeat sequence is cloned immediately downstream of the ATG that puts the downstream positioned β- galactosidase ORF out of frame. C(C) The identical transgenic array crossed into an *msh-6* genetic background.

Figure 5. *C.elegans* populations fed on *E.coli* that produce dsRNA homologues to the *C.elegans* genes *unc-22* (A) and *msh-6* (B).

Figure 6. Schematic representation of the high throughput RNAi based screens to identify novel mutator loci: Individual animals are fed on dsRNA producing bacteria, the progeny is collected and assayed for beta-galactosidase activity.

Table 1

Type of mutation	mutation	Position in *unc-93* ORF.	a.a. change
Frameshift +1	Insertion A	(221) TCGAGAA(A)TATTCGAA (235)	
+1	Insertion A	(229) ATTCGAAAAA(A)CTTCG (243)	
+1	Insertion A	(252) TTTGCAAAAA(A)TTTGG (266)	
+1	Insertion A	(252) TTTGCAAAAA(A)TTTGG (266)	
+1	Insertion A	(372) TTCCAAAAAA(A)GAAG (285)	
-1	Deletion T	(358) AAAGAGTTTTTCGAGG (373)	
Single basepair substitution.	G → A	(789) ATTTAACGGACTCCAA (804)	Gly → Arg
	G → A	(1155) ACACTGCGGACAAGTC (1170)	Gly → Arg
	G → A	(1551) TCTAGTTGGAGTTTAT (1566)	Gly → Arg
	G → A	(1650) TTCCCTAGTCTTCGGG (1665)	Val → Ile
	A → G	(1611) CTTTGTGATGGCCTGC (1626)	Met → Val
	A → C	(1492) AATATAAAGTTCATGT (1507)	Lys → Thr
	G → C	(1707) CGGAGCAGTAGTGAA (1721)	Val → Leu
	T → G	(1578) CGTCGGATGTGGCCTT (1593)	Cys → Gly
	T → G	Ggctctgaggtttcag AAAAATGGCT (1443)	Disruption of 3' splice site
Complex.	G→C +GC or +C ,G, +C	(67) AAAAGTAG(GC)ATCACCG (81) or (68) AAAGTA(C)G(C)ATCACCG (81)	
	TTTTTG ↓ CTTTTT	(523) GATCATTTTTGCCCGA (538) ↓ (523) GATCACTTTTTCCCGA (538)	His → His and Cys → Phe

Table 1. *Unc-93(e1500)* mutation spectrum in *C. elegans msh-6*

Table 2

	msh-6					Wild type	
Repeat	C36C5 $(A)_{15}$	F59A3 $(A)_{15}$	R03C1 $(A)_{15}$	C41D7 $(CA)_{18}$	M03F4 $(CA)_{18}$	F59A3 $(A)_{15}$	M03F4 $(CA)_{18}$
-1	0	3	2	7	5	0	0
0	44	42	38	32	34	44	44
+1	0	0	0	2	6	0	0
Total	44	45	40	41	45	44	44

Table 2: Microsatellite instability in the genome of *msh-6* mutants

Table 3

List of found mutants

Open reading frame	Similarity to known human genes
M04F3.1	Replication Protein A subunit 2 (rpa-2)
B0511.8	cdc-1
D1081.8	cdc-5
F02E9.4	sin-3
R06C7.7	
H26D21.2	msh-2
Y47G6A.11	msh-6
Y71F9AL.1/18	1 : N6 adenine-specific DNA methyl(transfer)ase, N12 18: Poly (ADP ribose) polymerase
F26E4.6	cytochrome c oxidase subunit VIIc
C01A2.3	cytochrome oxidase biogenesis protein like; OXA-1
F22D6.4	NADH ubiquinone oxidoreductase 13 kDa A subunit
F55A12.3	PI-4P5' kinase
E01A2.2	arsenate resistance protein 2 ARS-2
F25H2.9	proteasome zeta chain
C36B1.4	proteasome A type subunit
F39H11.5	proteasome beta chain

Table 4.

Gene name	Accession nr.	Similarity to known human genes
M04F3.1	NM_059045	Rpa2
B0511.8	NM_060382	Cdc1
D1081.8	NM_059902	Cdc5
F02E9.4	NM_059883	Sin3
R06C7.7	NM_059649	
H26D21.2	AF106587	Msh2
Y47G6A.11	AC024791	Msh6
Y71F9AL.18	NM_058671	Poly (adp ribose) polymerase
F55A12.3	AF003130	PI-4P5' kinase
E01A2.2	NM_058901	Arsenate resistance protein 2 (ars2)
F26E4.6	NM_060195	Cytochrome c oxidase su. VIIc
C01A2.3	NM_060955	Oxa1
F22D6.4	NM_059606	NADH ubiquinone oxidoreductase 13kDa su.
F25H2.9	NM_060364	Proteasome Z chain
C36B1.4	NM_059959	Proteasome A type su.
F39H11.5	Z81079	Proteasome beta chain
T02H6.11	NM_061394	Ubiquinol cyt. C reductase complex su.
F54D10.1	AF099917	Skr-15 SKP1 like
K07D4.3	AF077534	Rpn-11
C17G10.4	U28739	Cdc14
C25H3.3	NM_062714	
C25H3.4	NM_062713	Translation initiation factor SUI1
C32D5.6	NM_062872	
T19D12.5	NM_062948	Casein kinase I
B0495.2	NM_063216	Cdc2
F49E12.6	NM_063370	RBB3 like
T10B9.5	NM_063709	Cytochrome P450
R06F6.8a	Z46794	
R03D7.2	NM_063953	
F32A11.2	Z81521	Hpr-17 / rad-17
B0412.3	NM_064863	
R74.4	NM_065438	Heatshock protein
F20H11.5	NM_066052	D-amino acid oxidase
T26A5.5	U00043	
B0361.1		Cwf-19
H14A12.3	NM_066240	
T23G5.6	NM_066641	TdT interacting protein
Y56A3A.29	AL132860	Uracil-DNA glycosylase
T28D6.4	NM_067060	
Y111B2A.1	NM_067230	AFC2 like / CLK2-4 like

Table 4, continued

Y76A2B.5	NM_067400	
Y43F4B.1	NM_067336	
ZK520.3	NM_067423	
Y56A3A.33	NM_067164	Exonuclease similarity to antigen GOR
Y39A3CL.4a	AC024763	
Y62E10A.6	NM_070172	NADPH:adrenodoxin oxidoreductase
F29C4.6	NM_067464	
AC8.1	NM_075638	Poly (adp-ribose) polymerase
F15E6.1	NM_068138	
K08D10.2	NM_068105	
T05A12.4	NM_068659	
C33D9.5	NM_069115	Rad-50 like
K08F4.1	NM_069440	
K08E7.7	NM_070011	Cullin cul-6
K09B11.2	NM_070187	
F14F9.5	NM_071972	AP-endonuclease like
F44C4.4	NM_072280	Lin-15b like
ZC196.6	NM_072846	
ZK856.1	NM_073215	Cul-5 cullin
C06H2.3	NM_073430	
F08H9.4	NM_074185	Heatshock protein hsp20
F43D2.1	NM_074214	Cyclin C G1/S like
C30G7.1	NM_074279	Histone H1 like
C25D7.6	Z81079	MCM-3
F28E4.1		Cytochrome P450
Y113G7A.9	NM_075475	
W07A8.3	NM_75601	
F57C12.2	NM_075717	
F19G12.2	NM_075868	Ribonucleotide reductase
R07E4.2	NM_076596	SPT associated factor 42 like
C09B8.6	NM_076608	Heatshock protein hsp20
F45E1.6	NM_076943	Histone H3
C44C10.2	NM_077558	Cytochrome P450
F46G10.3	NM_077819	SIR2 family of genes
F02D10.7	NM_077840	
C53A5.3	Z81486	Hdac1
C35A5.9	NM_073298	Hdac2
H12C20.2a	AL022272	Pms-2
T28A8.7	Z92813	Mlh-1

References

Aaltonen et al., (1993) Science 260: 812-816

Baker et al., (1996) Nat. Genet. 13: 336-342

Brenner (1974) Genetics 77: 71-94

Bronner et al., (1994) Nature 368: 258-261

Brummelkamp et al, (2002) Sciencexpress 21 March 2002, 1-6

Cunningham et al., (1998) Cancer Res. 58: 3455-3460

de Wind et al., (1995) Cell 82: 321-330

Edelman et al., (1996) Cell 85: 1125-1134

Elbashir et al, (2002) Nature 411:494-498

Fishel et al., (1993) Cell 75: 1027-1038

Fire et al., (1998) Nature 391: 806-811

Fraser et al., (2000) Nature 408: 325-330

Herman et al., (1998) Proc. Natl. Acad. Sci. USA 95: 6870-6875

Ionov et al., (1993) Nature 260: 558-561

Jansen et al., (1999) Nat. Genet. 17: 119-121

Kane et al., (1997) Cancer Res. 57: 808-811

Kolodner et al., (1994) Cold Spring Harb Symp Quant Biol. 59: 331-338

Leach et al., (1993) Cell 75: 1215-1225

Liu et al., (1996) Nat. Med. 2: 169-174

Loeb (1991) Cancer Res. 51: 3075-3079

Lynch et al., (1998) Oncology 55: 103-108

Mello et al., (1991) EMBO J. 10: 3959-370

Narayanan et al., (1997) Proc. Natl. Acad. Sci. USA 94: 3122-3127

Nicolaides et al., (1994) Nature 371: 75-80

Papadopoulos et al., (1994) Science 268: 1915-1917

Peinado et al., (1992) Proc. Natl. Acad. Sci. USA 89: 10065-10069

Peltomaki et al., (1993) Cancer Res. 53: 5853-5855

Peltomaki and de la Chapelle (1997) Adv. Cancer Res. 71: 93-119

Prolla et al., (1998) Nat. Genet. 18: 276-279

Reitmar et al., (1995) Nat. Genet. 11: 64-70

Rosenbluth et al., (1983) Mut. Res. 110: 39-48

Tabara et al., (1999) Science 282: 430-431

Tijsterman, M., Pothof, J., and Plasterk, R.H.A. (2002). Frequent germline mutations and somatic repeat instability in DNA mismatch repair instability in DNA mismatch repair deficient *C. Elegans. Submitted*

Timmons and Fire (1999) Nature 395: 854

Veigl et al., (1998) Proc. Natl. Acad. Sci. USA 95: 8698-8702

Claims

1. A method for determining whether a product of a gene is involved in preventing a replication error in a cell comprising providing said cell with a specific inhibitor for said product and determining the level of functional expression of a marker gene in said cell, wherein the level of expression of said marker gene is dependent on the occurrence of said replication error.
2. A method according to claim 1, wherein said replication error comprises nucleic acid repeat instability.
3. A method according to claim 1 or claim 2, wherein said specific inhibitor for said product comprises gene specific RNA.
4. A method according to claim 3, wherein said specific inhibitor for said product comprises gene specific double stranded RNA.
5. A method according to any one of claims 1-4, wherein said cell is present in a non-human organism.
6. A method according to claim 5, wherein said organism comprises *C. elegans*.
7. A method according to any one of claims 1-6, wherein said marker gene is provided to said cell.
8. A method according to claim 7, wherein said marker gene comprises LacZ.
9. A method according to any one of claims 1-8, wherein the level of expression of said marker gene is dependent on a nucleic acid repeat in said gene.
10. A method according to any one of claims 1-9, wherein said expression of said marker gene is dependent on said nucleic acid repeat because said repeat, or an incorrect repair of said repeat results in a frame shift within the coding region of said marker gene.
11. A method according to claim 10, wherein said frame shift results in a functional protein.

12. A method according to claim 11, wherein an activity of said functional protein is detected.

13. A method according to claim 12, wherein said activity comprises β-galactosidase activity.

14. A method according to any one of claims 1-13, further comprising identifying said gene involved in preventing nucleic acid repeat instability in a cell.

15. An isolated and/or recombinant gene obtainable by a method according to claim 14.

16. An isolated and/or recombinant gene according to claim 15, wherein said gene is a gene as listed in table 3 or table 4 of this application.

17. A method for determining whether a cell is predisposed to display a nucleic acid repeat instability phenotype comprising determining functional expression of a gene according to claim 15 or claim 16, or an equivalent or homologue thereof, in said cell.

18. A method according to claim 17, wherein said cell is part of a non-human organism.

19. A method according to claim 17, wherein said cell is present in a clinical sample.

20. A method according to claim 19, further comprising determining whether an individual is predisposed to display a nucleic acid repeat instability phenotype.

21. A method according to claim 19 or claim 20, further comprising determining whether said cell is a cancer cell.

22. A kit for performing a method according to any one of claims 17-21, comprising a means for determining functional expression of a gene identifiable with a method according to claim 14 or of a gene listed in table 3 or table 4 or an equivalent or homologue thereof.

23. A kit according to claim 22, comprising an antibody specific for a gene product of a gene identifiable with a method according to claim 14 or of a gene listed in table 3 or table 4, or an equivalent or homologue thereof.
24. A kit according to claim 22 or claim 23, comprising a probe for a gene identifiable with a method according to claim 14 or of a gene listed in table 3 or table 4., or an equivalent or homologue or a gene product thereof.
25. A kit according to any one of claims 22-24, comprising means for obtaining a sequence of a gene identifiable with a method according to claim 14 or of a gene listed in table 3 or table 4., or an equivalent or homologue thereof ,or a sequence of a gene product thereof.
26. A method for determining whether a compound is capable of influencing a process involved in preventing a replication error in a cell comprising providing said cell with said compound and determining the level of expression of a marker gene in said cell, wherein the level of expression of said marker gene is dependent on said replication error.
27. A method according to claim 26, further comprising providing said cell with a specific inhibitor for the expression of a gene involved in preventing a replication error in a cell.
28. A method according to claim 27, wherein said gene involved in preventing a replication error in a cell, is a gene obtainable with a method according to any one of claims 1-14, or a gene listed in table 3 or table 4, or an equivalent or homologue thereof.
29. A gene delivery vehicle comprising a nucleic acid according to claim 15 or claim 16.
30. A method for influencing a process involved in preventing a replication error in a cell comprising providing said cell with a gene delivery vehicle according to claim 29.
31. Use of a gene delivery vehicle according to claim 29 for the preparation of a medicament.

32. A non-human animal comprising a marker gene wherein the level of expression of said marker gene is dependent on the occurrence of said replication error.

33. A non-human animal according to claim 32 wherein said marker gene is provided to cells of said animal.

34. A non-human animal according to claim 32 or claim 33, wherein said animal is transgenic for said marker gene.

35. A method for determining whether a compound is capable of inducing a replication error comprising providing a non-human animal according to any one of claims 32-34, with said compound and determining in said animal or progeny thereof whether the expression level of said marker gene is altered.

36. A method according to claim 35, wherein said non-human animal comprises *C. elegans*.

37. A method according to claim 36, wherein said compound comprises RNAi, or a free radical.

38. A method according to claim 37, wherein said RNAi is specific for a gene listed in table 3 or table 4.

39. A method for typing a cell comprising determining in a sample comprising said cell functional expression of a gene listed in table 3 or table 4, or an equivalent or a homologue thereof and comparing said functional expression with a reference sample.

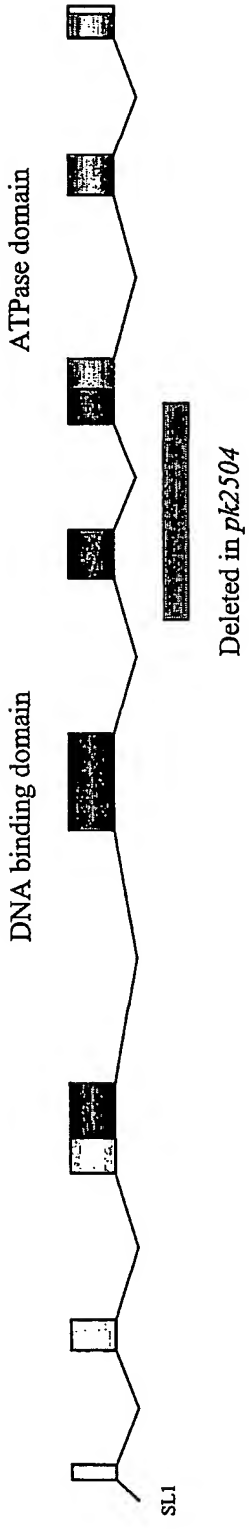
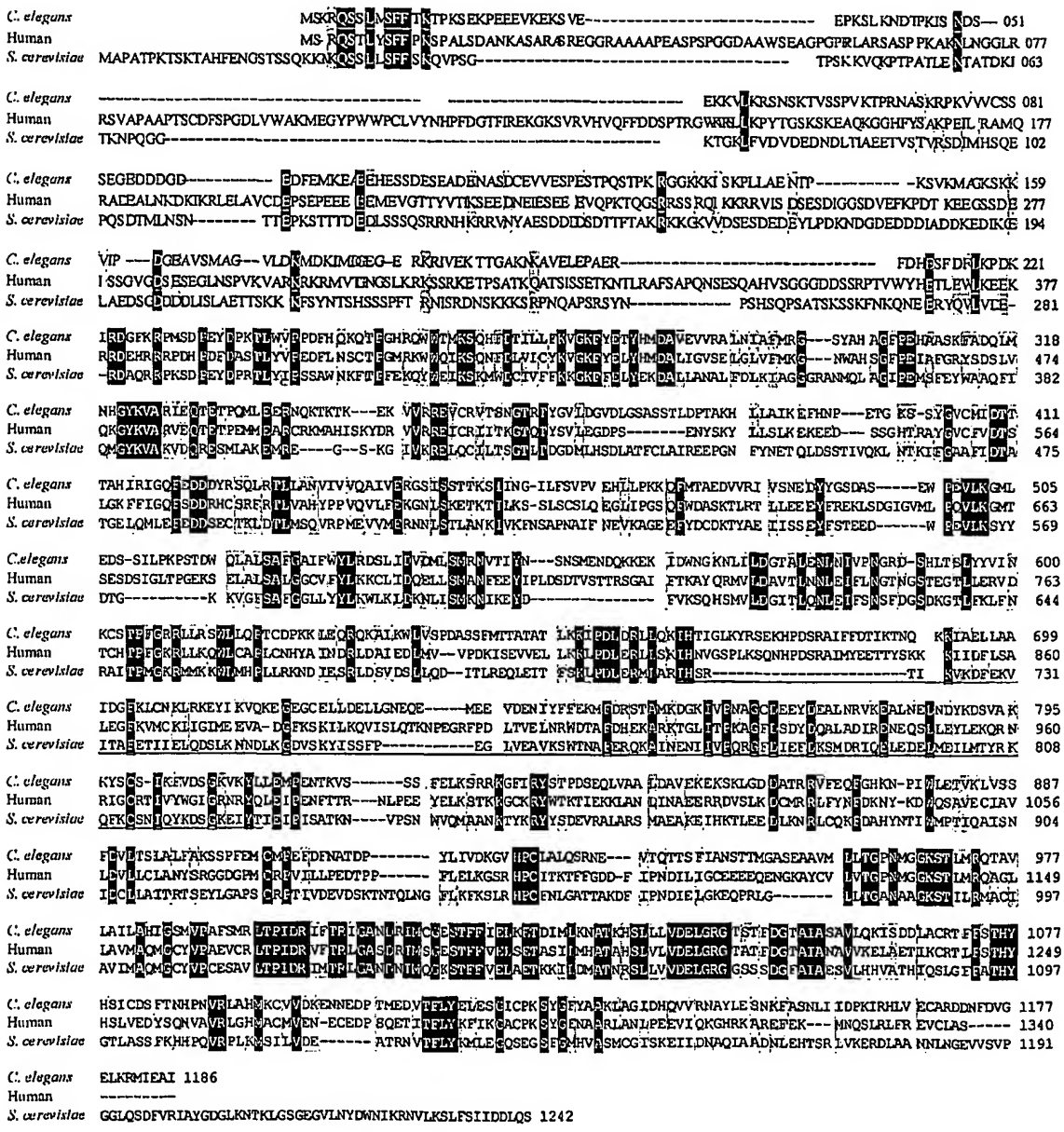
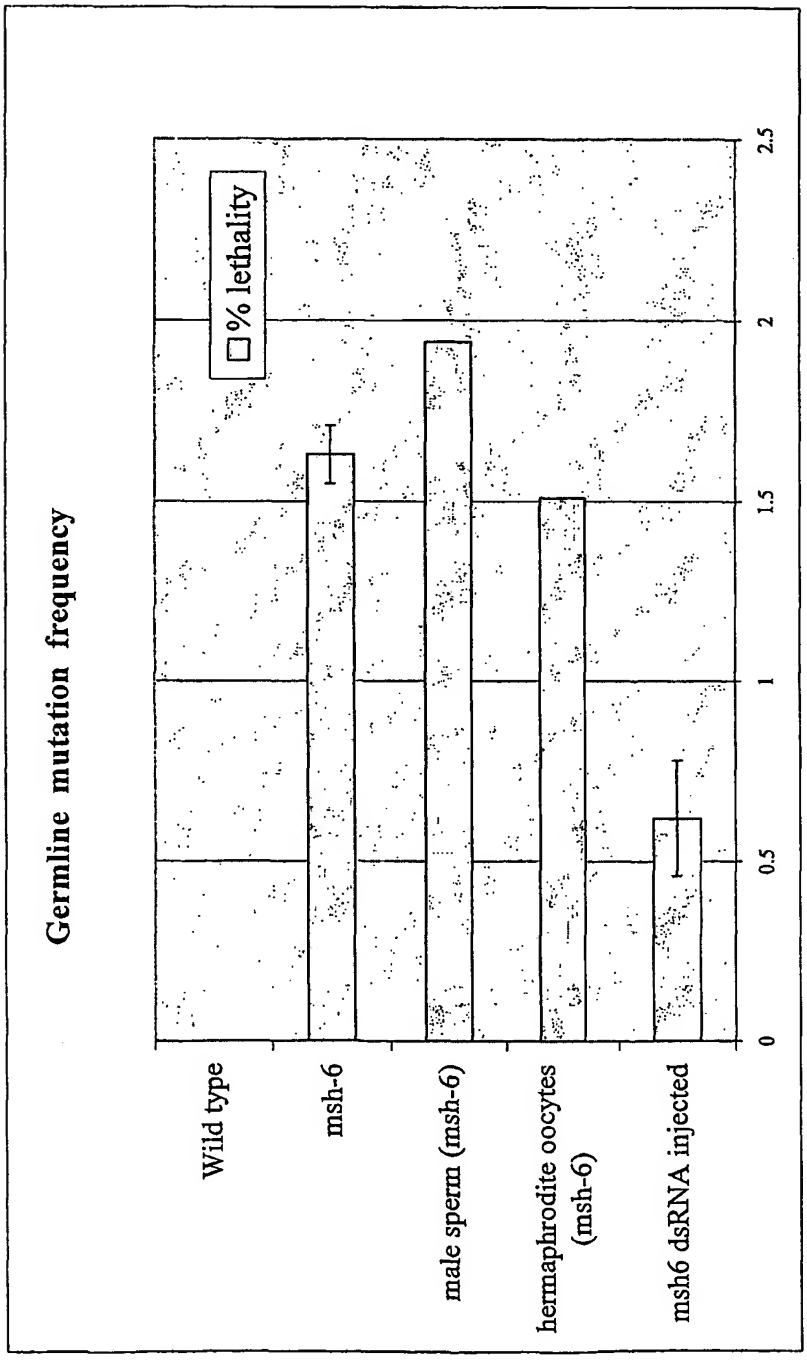


Fig. 1A

Fig. 1B





Spontaneous germline mutation frequency in wildtype *C. elegans*, *msh-6* genetic mutants and wildtype *C. elegans* exposed to *msh-6* dsRNA.

Fig. 2

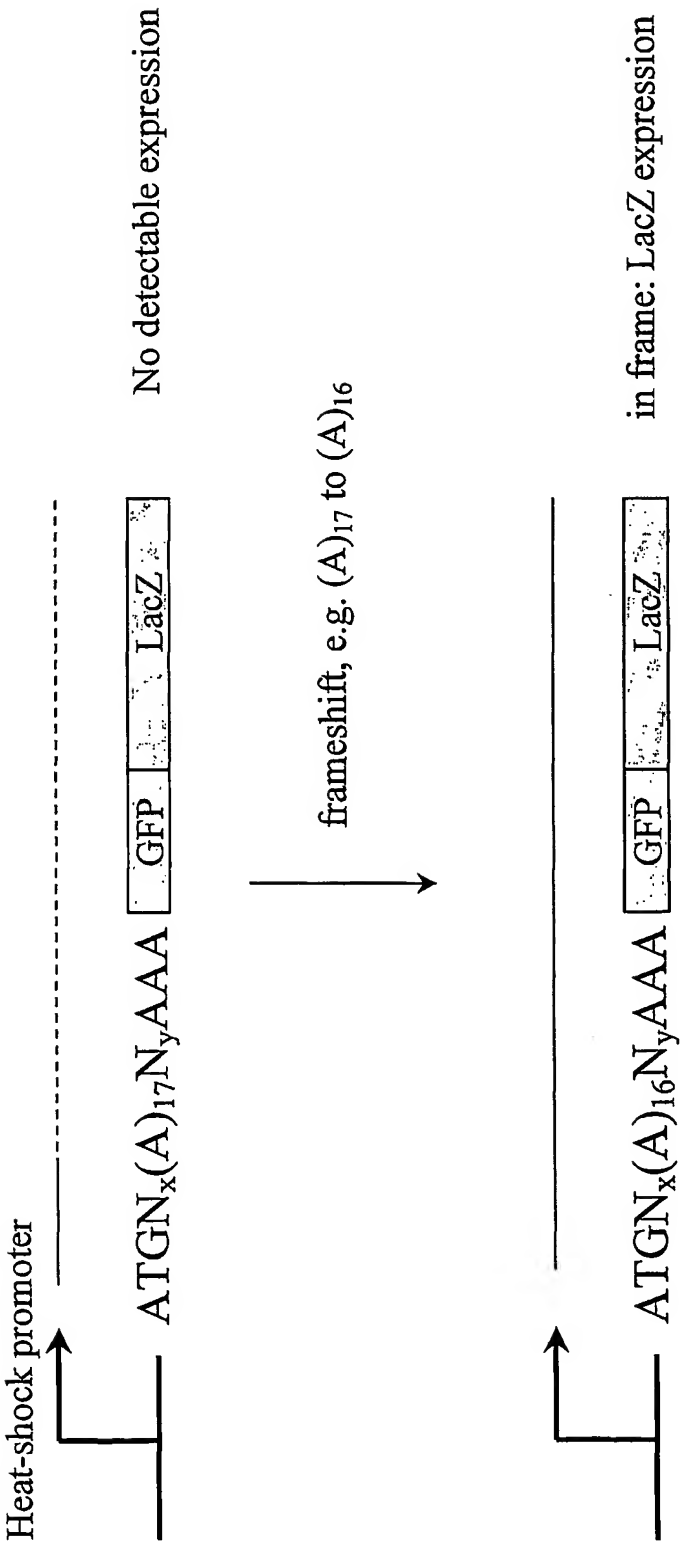


Fig. 3: Outline of the principle to detect somatic repeat instability

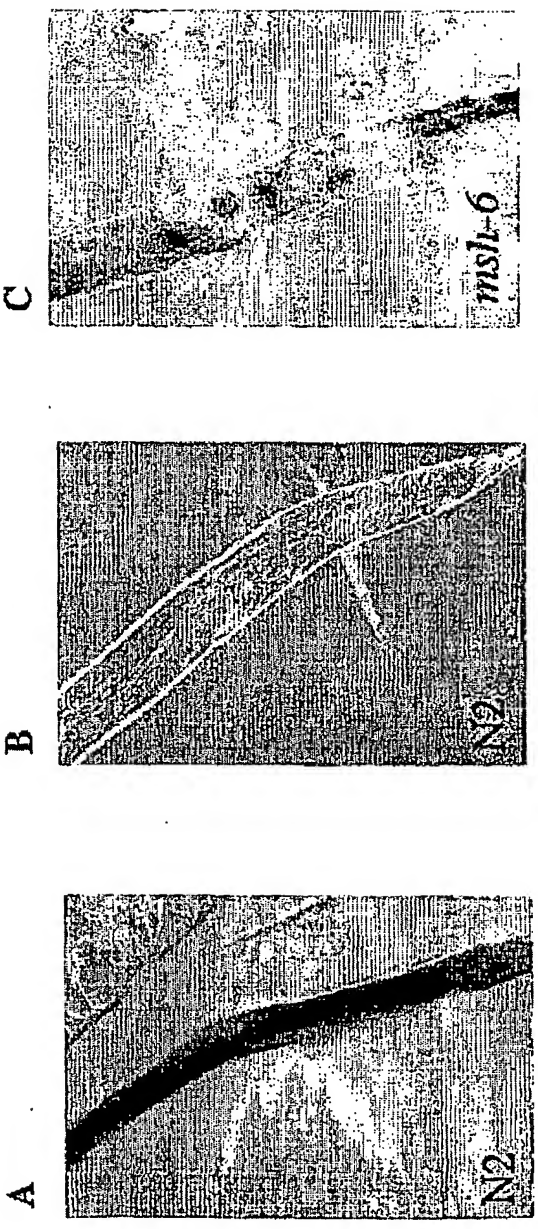


Figure 4: A) Wildtype *C.elegans* containing the in-frame construct, B) the +1 out of frame construct. C) Genetic *msh-6* mutants that contain the +1 out of frame construct display LacZ expression.

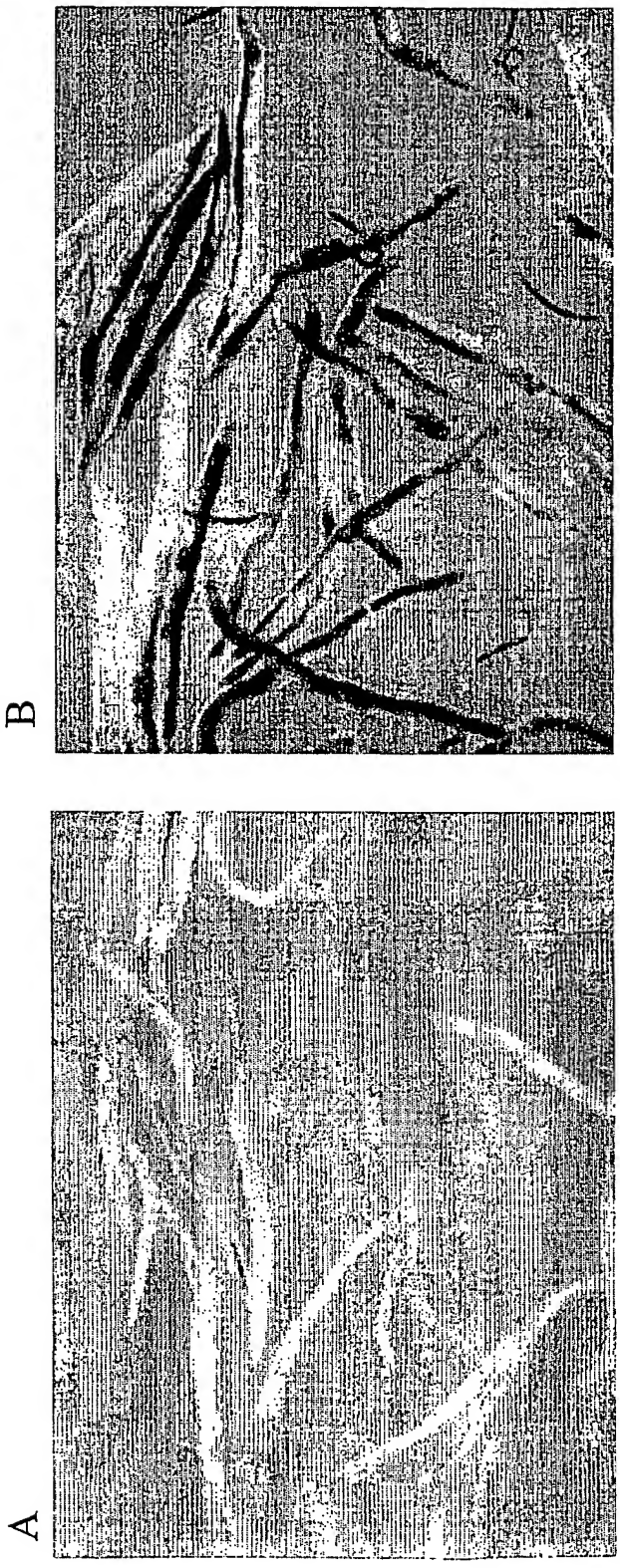


Fig 5: *C. elegans* populations fed on *E.coli* that produce dsRNA homologues to the *C. elegans* genes *unc-22* (A) and *msh-6* (B)

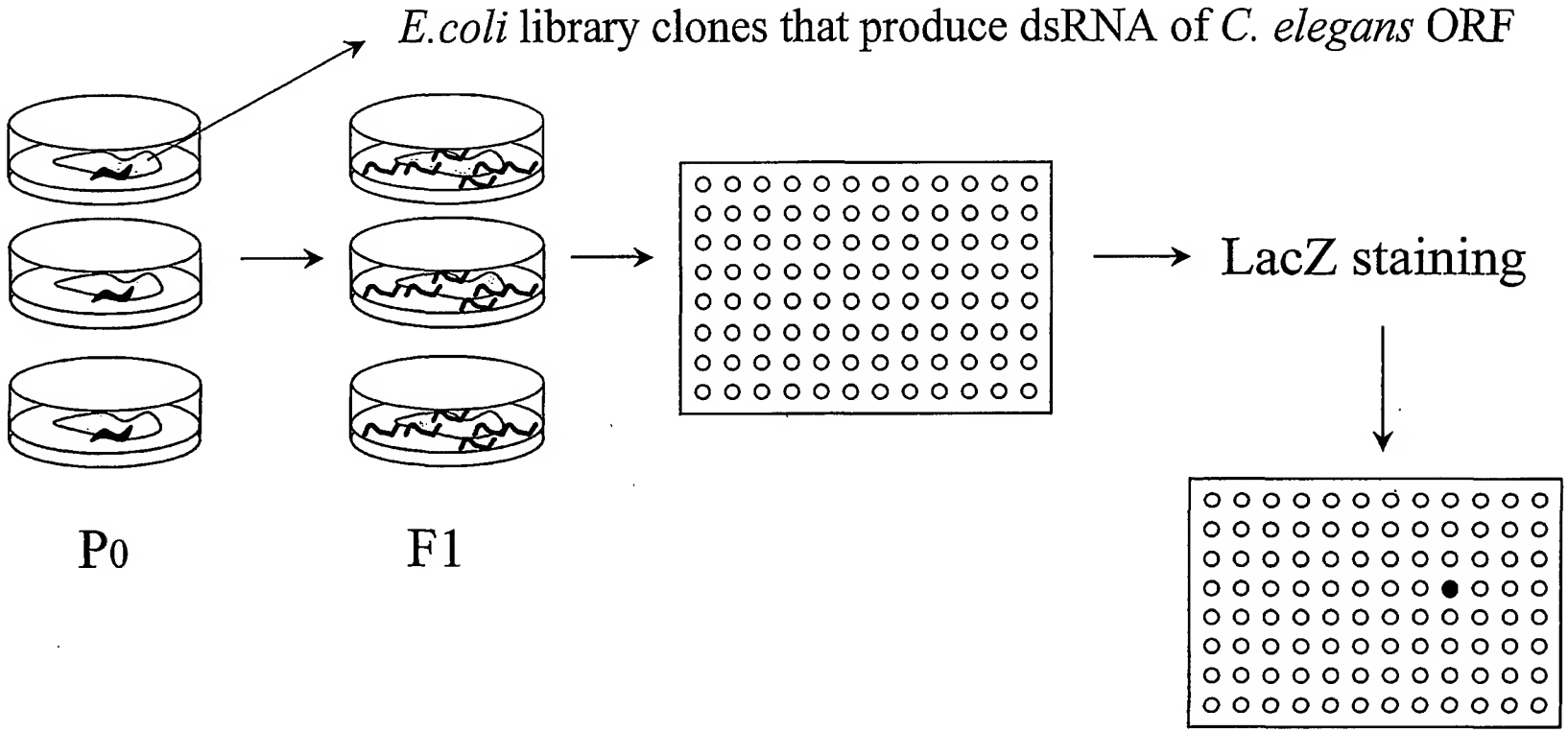


Fig. 6: Schematic representation of the high throughput RNAi based screens to identify novel mutator loci: Individual animals are fed on dsRNA producing bacteria, the progeny is collected and assayed for beta-galactosidase activity.